*Human Molecular Genetics, 2002, Vol. 11, No. 24* **3031–3038**

# Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels

**Len A. Pennacchio[1,*,†], Michael Olivier[2,†], Jaroslav A. Hubacek[3,†], Ronald M. Krauss[1], Edward M. Rubin[1] and Jonathan C. Cohen[3]**

[1]Genome Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [2]Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI, USA and [3]Center for Human Nutrition and McDermott Center for Human Growth and Development, UT Southwestern Medical Center, Dallas, TX, USA

Received July 15, 2002; Revised and Accepted September 16, 2002

**The recently identified apolipoprotein A5 gene (*APOA5*) has been shown to play an important role in determining plasma triglyceride concentrations in humans and mice. We previously identified an *APOA5* haplotype (designated *APOA5*2*) that is present in ~16% of Caucasians and is associated with increased plasma triglyceride concentrations. In this report we describe another *APOA5* haplotype (*APOA5*3*) containing the rare allele of the single nucleotide polymorphism c.56C>G that changes serine to tryptophan at codon 19 and is independently associated with high plasma triglyceride levels in three different populations. In a sample of 264 Caucasian men and women with plasma triglyceride concentrations above the 90th percentile or below the 10th percentile, the *APOA5*3* haplotype was more than three-fold more common in the group with high plasma triglyceride levels. In a second independently ascertained sample of Caucasian men and women ($n = 419$) who were studied while consuming their self-selected diets as well as after high-carbohydrate diets and high-fat diets, the *APOA5*3* haplotype was associated with increased plasma triglyceride levels on all three dietary regimens. In a third population comprising 2660 randomly selected individuals, the *APOA5*3* haplotype was found in 12% of Caucasians, 14% of African-Americans and 28% of Hispanics and was associated with increased plasma triglyceride levels in both men and women in each ethnic group. These findings establish that the *APOA5* locus contributes significantly to inter-individual variation in plasma triglyceride levels in humans. Together, the *APOA5*2* and *APOA5*3* haplotypes are found in 25–50% of African-Americans, Hispanics and Caucasians and support the contribution of common human variation to quantitative phenotypes in the general population.**

## INTRODUCTION

Increased plasma triglyceride levels are associated with atherosclerotic heart disease, the leading cause of death in the United States (1,2). Plasma triglyceride levels vary widely both among and within individuals and environmental factors that confer coronary risk, such as cigarette smoking, obesity and lack of exercise are frequently associated with elevated plasma triglyceride concentrations (3,4). Data from nuclear families (5–7) and twin studies (8,9) indicate that triglyceride levels are also strongly influenced by genetic factors, although the heritability estimates vary widely (20–80%) among different studies. While heritability estimates based on nuclear families are potentially confounded by the fact that relatives who live together share environment as well as genes, the similarity of plasma triglyceride levels in identical twins reared apart provides compelling evidence that genetic polymorphism is a major cause of inter-individual variation in plasma triglyceride concentrations (9,10).

Several studies have attempted to identify specific genetic determinants of plasma triglyceride concentrations. Since triglyceride concentrations do not segregate in a clearly Mendelian fashion in most families, the majority of studies have sought statistical associations between DNA sequence polymorphisms in candidate genes and plasma triglyceride concentrations in cohorts of unrelated patients. Most of the associations reported have been modest and difficult to reproduce (11,12). The most consistent association observed to date has linked a rare allele at the *APOA1/C3/A4* locus to increased plasma triglyceride concentrations (13). The allele, defined by a G to C substitution in the 3′ untranslated region of the *APOC3* gene that interrupts a recognition site for the

*To whom correspondence should be addressed at: Department of Genome Sciences, MS 84-171, One Cyclotron Road, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA. Tel: +1 5104867498; Fax: +1 5104864229; Email: lapennacchio@lbl.gov

†The authors wish it to be known that, in their opinion, these three authors should be considered as joint First Authors.

restriction enzyme *Sst*I, has been associated with hypertriglyceridemia in a variety of different ethnic groups (14–24).

More recently plasma triglyceride concentrations were found to be strongly associated with polymorphisms in a newly identified apolipoprotein gene, *APOA5*. *APOA5* was discovered when 200 kb of orthologous sequence spanning the *APOA1/C3/A4* gene cluster were compared in humans and mice (25). Studies in mice that were genetically modified to either overexpress or lack *apoA5* provided direct evidence that the protein product plays a role in plasma triglyceride metabolism (25). Plasma triglyceride concentrations were decreased by ~66% in mice overexpressing a human *APOA5* transgene and were increased four-fold in mice in lacking *apoA5*. To determine if sequence variation in *APOA5* contributed to differences in plasma triglyceride levels in humans, association studies were performed. A minor haplotype of *APOA5* (*APOA5*2*, Fig. 1B) defined by three polymorphisms (c.1259T>C, IVS3 + 476G>A and −1131T>C, designated SNP1, SNP2 and SNP3, respectively) was associated with a 20–30% elevation in plasma triglyceride levels in 500 unrelated Caucasian men and women (Berkeley Lipid Study Population) (25). Analysis of the *APOC3 Sst*I polymorphism in these individuals indicated that the association was not due to effects of this previously defined neighboring SNP (25). The association was confirmed in an independently ascertained cohort of Caucasian men and women. The *APOA5*2* haplotype was more than three-times as common in individuals who had plasma triglyceride concentrations greater than the 90th percentile than in those with plasma triglyceride levels below the 10th percentile for age and sex (stratified population).

Taken together, these data indicate that APOAV plays an important role in plasma triglyceride homeostasis and that polymorphisms in *APOA5* are associated with variation in plasma triglyceride levels in humans. The present study was undertaken to further define the relationship between sequence variations in *APOA5* and plasma triglyceride levels in humans. DNA sequencing was used to screen the coding region and proximal promoter of *APOA5* for DNA sequence variations in a large number of hypertriglyceridemic individuals and the polymorphisms identified were tested for association with plasma triglyceride levels in three independently ascertained groups of individuals. Our data indicate that 25–50% of individuals in three major ethnic groups (African-Americans, Hispanics and Caucasians) carry at least one of two *APOA5* haplotypes that are independently associated with elevated plasma triglyceride concentrations. These findings support the theory that common genetic variation in the general population significantly contributes to quantitative phenotypes in humans.

## RESULTS

### DNA sequencing

Screening of the coding regions and intron–exon boundaries of *APOA5* in 116 hyperlipidemic individuals revealed 9 new DNA sequence variations (Fig. 1A). An A to G substitution 3 nucleotides upstream of the initiation codon (c.−3A>G) was found to be in strong linkage disequilibrium with three previously described polymorphisms (c.1259T>C, IVS3 + 476G>A and −1131T>C, Fig. 1B) that define the *APOA5*2* haplotype previously associated with increased plasma triglyceride concentrations (25). The A to G substitution results in a conservative change at position −3 in the predicted translation initiation consensus sequence (26,27). A common nonsynonymous substitution was also identified which results in a C to G substitution (c.56C>G) changing codon 19 from serine to tryptophan in 23 individuals. A second nonsynonymous substitution (c.944C>T) that changed codon 315 from alanine to valine was identified in two hyperlipidemic individuals. This conservative substitution did not co-segregate with hyperlipidemia in the family members of one of these individuals (data not shown) and was not found in 108 normolipidemic individuals, therefore no further studies of this polymorphism were undertaken. The other six polymorphisms, including three silent substitutions (c.132C>A, c.695C>G, c.738C>T) and three polymorphisms each found only in single individuals (IVS2 +55G>C and c.1132C>T and c.1156 G>A in the 3′ UTR) were not evaluated further.

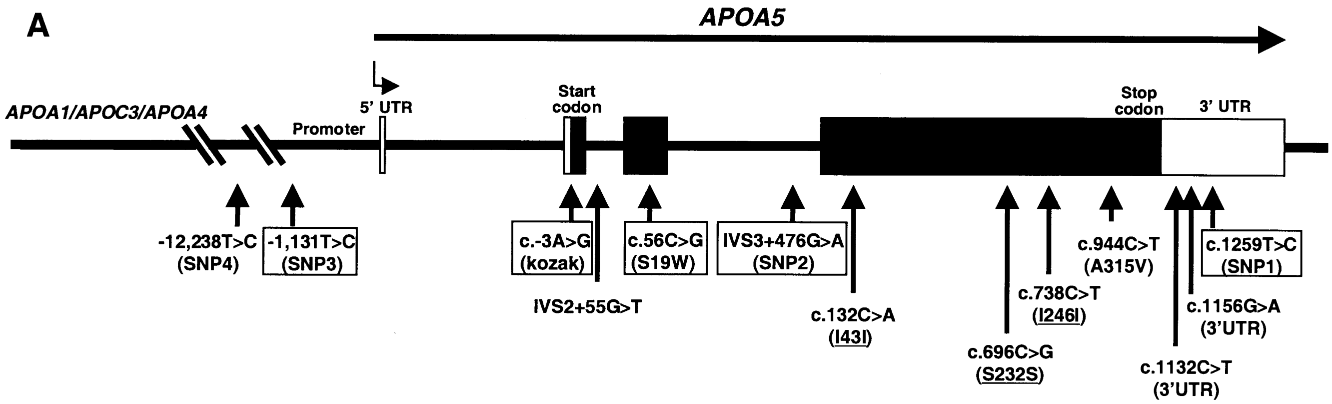
### Allele frequency and linkage disequilibrium

Five polymorphisms were found to define three common haplotypes (denoted *APOA5*1*, APOA5*2, and *APOA5*3*) in 419 unrelated Caucasian individuals (Berkeley Lipid Study Population; these samples were previously described in ref. 25; Fig. 1B). These three haplotypes represented 82%, 8% and 8% of the *APOA5* chromosomes examined and thus comprise ~98% of *APOA5* haplotypes in this population. *APOA5*2* is distinguished from the common haplotype (*APOA5*1*) by four nucleotide substitutions (−1131T>C, c.−3A>G, IVS3+ 476G>T and c.1259T>C) and was previously shown to be associated with increased plasma triglyceride levels (25). *APOA5*3* is distinguished from the common haplotype by the substitution of G for C at nucleotide c.56 (codon 19 in the amino acid sequence). To determine the relative frequencies of the −1131T>C polymorphism in African-Americans and Hispanics, this variant was assayed in the DHDPP population. The allele frequency was significantly higher in African-Americans (0.12) and Hispanics (0.16), than in Caucasians (0.06, $P < 0.001$). The frequency of the W19 allele (which defines haplotype *APOA5*3* in Caucasians) was similar in African-Americans (0.07) and Caucasians (0.06), but was substantially higher in Hispanics (0.15, $P < 0.001$ compared to African-Americans).

### Association studies

To test for association between the common nonsynonymous polymorphism identified in this study (S19W) and plasma triglyceride concentrations, the allele frequencies at this locus were compared in Caucasian men and women who had plasma triglyceride concentrations above the 90th percentile or below the 10th percentile for age and sex (stratified population). To eliminate confounding by the *APOA5*2* haplotype that was previously associated with high plasma triglyceride levels (25), individuals who carried this haplotype were excluded. In both sexes, the rare allele at codon 19 (W19) was significantly more common in individuals with plasma triglyceride levels above the 90th percentile than in those with plasma triglyceride levels below the 10th percentile (Table 1). Since individuals with the

**A**




**B**

| Haplotype | Freq. | -1131T>C (SNP3) | c.-3A>G | c.56C>G (S19W) | IVS3+476G>A (SNP2) | c.1259T>C (SNP1) |
|---|---|---|---|---|---|---|
| ***APOA5*1*** | 81.6% | T | A | C | G | T |
| ***APOA5*2*** | 8.0 | **C** | **G** | C | **A** | **C** |
| ***APOA5*3*** | 8.0 | T | A | **G** | G | T |

**Figure 1.** (**A**) *APOA5* genomic structure and polymorphism location. The gene is transcribed from left to right as indicated by the large horizontal arrow. Exons are depicted by boxes with protein-encoding regions shaded black. The position and identity of SNPs identified in *APOA5* are shown below the schematic. SNPs found within the open reading frame show the predicted amino acid substitution in parentheses (synonymous changes are underlined). SNPs previously identified are indicated as SNPs1–4 in parentheses (25). For all the SNPs the major allele basepair sequence is identical to that found in chimpanzee, except for −12,238T>C where they are reversed. (**B**) Common *APOA5* haplotypes and their relative frequencies in 419 Caucasian samples (Berkeley Lipid Study Population). The five SNPs used in this analysis are boxed in panel A. The SNPs are depicted in the following order: −1131T>C, c.−3A>G, c.56C>G, IVS3 + 476G>A, c.1259T>C. Haplotype frequencies were predicted using the Expectation-Maximization algorithm (5000 iterations, (31)). The three depicted haplotypes account for 97.6% of all haplotypes, none of the other predicted haplotypes had a frequency of greater than 1%. The minor alleles that define the haplotypes are highlighted in bold. In an independent Caucasian population, we find the *APOA5*1, APOA5*2* and *APOA5*3* haplotype frequencies are 83.4%, 8.0% and 8.4%, respectively. These samples are from 367 unrelated Caucasian individuals of Northern European descent, collected in the Midwest.

**Table 1.** *APOA5* genotypes in men and women with high (>90th percentile) and low (<10th percentile) plasma triglyceride concentrations (stratified population)

| | *APOA5* genotype | | | *P* |
|---|---|---|---|---|
| | S19/S19 | S19/W19 | W19/W19 | |
| Men TG < 10th percentile ($n = 82$) | 74 (90.5) | 7 (8.5) | 1 (1) | <0.005 |
| Men TG > 90th percentile ($n = 82$) | 63 (77) | 19 (23) | 0 (0) | |
| Women TG < 10th percentile ($n = 50$) | 50 (100) | 0 (0) | 0 (0) | <0.001 |
| Women TG > 90th percentile ($n = 50$) | 39 (78) | 11 (22) | 0 (0) | |

Values are numbers of individuals in each group. The percentage of individuals with the genotype is given in parentheses. All individuals were homozygous for the −1131T allele that defines the *APOA5*2* haplotype. *P* values were calculated using Fisher's exact test.

−1131C allele were excluded, the association between the S19W polymorphism and plasma triglyceride concentrations is independent of the *APOA5*2* haplotype that was previously shown to be associated with increased plasma triglyceride levels (25).

To further assess this association, the S19W polymorphism was assayed in 419 healthy independently-ascertained Caucasians (354 men and 65 women) (Berkeley Lipid Study Population). Baseline blood samples were obtained from these individuals on their self-selected diets and additional samples were drawn following the consumption of a defined high-carbohydrate or high-fat diet (a further description of these samples can be found in ref. 25). On all three diets, individuals

**Table 2.** *APOA5* haplotypes and plasma lipid and lipoproteins in men and women consuming normal, high-fat or low-fat diets (Berkeley Lipid Study Population)

| | *APOA5* haplotype | | | | ANOVA |
|---|---|---|---|---|---|
| | *APOA5*1/*1* | *APOA5*1/*2* | *APOA5*1/*3* | *APOA5*2/*3* | *P* |
| *n* | 308 | 66 | 40 | 5 | |
| Triglyceride | | | | | |
| Baseline | $112.6 \pm 3.8$ | $147.5 \pm 11.6$ | $150.8 \pm 14.0$ | $208.6 \pm 50.0$ | <0.0001 |
| High fat | $92.7 \pm 3.0$ | $130.7 \pm 12.9$ | $127.5 \pm 12.1$ | $160.4 \pm 32.5$ | <0.0001 |
| Low fat | $122.0 \pm 4.5$ | $151.8 \pm 12.7$ | $167.3 \pm 19.3$ | $220.8 \pm 56.6$ | 0.0004 |
| VLDL mass | | | | | |
| High fat | $65.7 \pm 3.2$ | $110.9 \pm 12.6$ | $107.5 \pm 12.4$ | $140.2 \pm 34.7$ | <0.0001 |
| Low fat | $110.1 \pm 5.2$ | $137.9 \pm 12.8$ | $167.0 \pm 17.1$ | $224.6 \pm 66.0$ | 0.0001 |
| ApoB | | | | | |
| High fat | $90.7 \pm 1.4$ | $97.4 \pm 2.8$ | $109.4 \pm 3.8$ | $99.4 \pm 13.7$ | <0.0001 |
| Low fat | $93.0 \pm 1.4$ | $100.2 \pm 3.0$ | $106.8 \pm 3.6$ | $116.0 \pm 4.9$ | 0.0005 |

*APOA5* haplotypes **1*, **2*, and **3* are defined by the two SNPs S19W and −1131T>C (see Fig. 1B). Differences among genotypes were analysed by one way ANOVA using STATVIEW 4.1 software (Abacus Concepts, Inc, Berkeley, CA). Values are given in mg/dl ± SEM. Three individuals were homozygous for *APOA5*2*. Their plasma triglyceride levels were 147, 386 and 95 mg/dl. Two individuals were homozygous for *APOA5*3*. Their plasma triglyceride levels were 127 and 250 mg/dl. VLDL mass and ApoB levels were not determined under baseline dietary conditions in this population.

who were heterozygous for the W19 allele and who lacked haplotype *APOA5*2* had significantly higher plasma triglyceride concentrations than did individuals homozygous for the S19 allele (Table 2). The increase in mean plasma triglyceride levels associated with a single copy of the W19 allele was ~36%, which is similar to the increase in triglyceride levels associated with *APOA5*2* haplotype (~32%) in other members of this study (25). The five individuals containing both the rare −1131C and W19 alleles (which define *APOA5*2* and *APOA5*3*, respectively) had an increase in mean plasma triglycerides of 85% compared to individuals homozygous for both the −1131T and S19 alleles (*APOA5*1* haplotype).

To determine if the W19 allele was associated with increased plasma triglyceride concentrations in other ethnic groups, the S19W polymorphism was assayed in a random sample of 1392 African-American, 420 Hispanic and 848 Caucasians (DHDPP Population). In both sexes of all three ethnic groups, both the mean and the median plasma triglyceride concentrations were higher in W19 heterozygotes than in S19 homozygotes (Table 3). The difference was significant at the 0.05 confidence level for African-Americans and Caucasians in both sexes, but did not achieve the nominal significance threshold in Hispanics ($P = 0.087$ and 0.057 for men and women, respectively), presumably due to the smaller sample size in this group. In this population, the previously described −1131C allele was associated with increased plasma triglyceride concentrations in Hispanic men and women and in Caucasian men, but not in African-American men and women or in Caucasian women (Table 4) (25).

To determine the relative risk of hypertriglyceridemia conferred by the W19 allele, the proportion of individuals with each genotype who had plasma triglycerides exceeding the 90th percentile for their sex and race was calculated in 2660 randomly selected individuals. The proportion of individuals with plasma triglycerides exceeding the 90th percentile increased from 9% of SS homozygotes to 15% of SW heterozygotes and 37% of WW homozygotes, giving odds ratios of 1.8 (95% confidence interval 1.3 to 2.6, $P < 0.001$) for SS versus SW and 5.6 (95% confidence interval 2 to 15, $P < 0.001$) for SS versus WW, respectively. The risk of high plasma triglyceride levels (>90th percentile) was significantly greater in WW homozygotes than in SW heterozygotes ($P < 0.04$, Fisher's exact test). Essentially identical results were obtained when individuals with the −1331C allele were excluded. Among individuals who were homozygous for the S19 allele, the odds ratio for high plasma triglycerides was not increased among −1131TC heterozygotes (odds ratio 1.2 with respect to −1311TT, $P > 0.3$), but was significantly increased among −1131CC homozygotes (odds ratio 4.5, $P < 0.005$). The odds ratio for hypertriglyceridemia was higher among individuals carrying both the W19 and the −1131C alleles than among individuals carrying only one of these alleles (1.77, 95% confidence interval 0.83–3.76, $P = 0.14$) although the difference did not achieve the 0.05 significance level, presumably because of the small number of compound heterozygotes.

## DISCUSSION

Plasma triglyceride concentrations reflect a complex interaction between multiple genetic and environmental factors. To further define the genetic factors that predispose to elevated plasma triglyceride concentrations, we screened the *APOA5* gene in individuals with hypertriglyceridemia. DNA sequencing revealed 9 new sequence variations, including a C to G substitution (c.56C>G) that changed codon 19 from serine to tryptophan. The W19 allele defines a new haplotype (*APOA5*3*) that was significantly more common among men and women with high plasma triglyceride concentrations (>90th percentile for age and sex) (stratified population) and was systematically associated with increased plasma triglyceride concentrations in men and women from three different ethnic groups (African-American, Hispanic and Caucasian) (DHDPP Population). The effect of the *APOA5*3* haplotype on plasma triglyceride levels was also evident in individuals

**Table 3.** Ethnicity, *APOA5* genotype and plasma triglyceride concentrations in a random sample of Dallas County residents (DHDPP Population)

| | | S19/S19 | S19/W19 | W19/W19 | *P* |
|---|---|---|---|---|---|
| African-American women | Mean ± SD | 101 ± 169 | 131 ± 120 | 141 ± 50 | 0.0023 |
| | Median ± IQ range | 80 ± 59 | 97 ± 85 | 192 | |
| | *n* | (707) | (108) | (6) | |
| African-American men | Mean ± SD | 132 ± 152 | 176 ± 319 | 264, 84 | 0.024 |
| | Median ± IQ range | 94 ± 84 | 111 ± 92 | | |
| | *n* | (494) | (75) | (2) | |
| Hispanic women | Mean ± SD | 143 ± 95 | 174 ± 209 | 394 ± 534 | 0.057 |
| | Median ± IQ range | 119 ± 92 | 135 ± 99 | 214 | |
| | *n* | (185) | (57) | (7) | |
| Hispanic men | Mean ± SD | 173 ± 139 | 204 ± 182 | 206, 124 | 0.087 |
| | Median ± IQ range | 139 ± 108 | 157 ± 101 | | |
| | *n* | (119) | (50) | (2) | |
| Caucasian women | Mean ± SD | 124 ± 96 | 147 ± 90 | 237, 125 | 0.012 |
| | Median ± IQ range | 100 ± 87 | 122 ± 110 | | |
| | *n* | (386) | (54) | (2) | |
| Caucasian men | Mean ± SD | 161 ± 121 | 255 ± 225 | | 0.0012 |
| | Median ± IQ range | 126 ± 116 | 183 ± 237 | | |
| | *n* | (362) | (44) | (0) | |

Values are plasma triglyceride concentrations (mg/dl). For groups represented by only two individuals, the individual plasma triglyceride concentrations are given. One-sided *P*-values were calculated using Wilcoxon's test.

consuming three different dietary regimens and was independent of the *APOA5∗2* haplotype that was previously associated with increased plasma triglyceride levels (Berkeley Lipid Study Population) (25). Taken together, these data provide compelling evidence that polymorphisms in the *APOA5* gene confer heritable variation in plasma triglyceride levels and firmly establish a role for APOAV in human triglyceride metabolism.

All 9 of the DNA sequence variations identified in this study were single nucleotide substitutions. Seven of these, including two rare noncoding variants, four silent substitutions, and a rare nonsynonymous substitution that did not segregate with elevated plasma triglyceride concentrations in a nuclear family, are unlikely to be important sources of inter-individual variation in plasma triglyceride levels. The other two polymorphisms (−3c.A>G and S19W) were both associated with increased plasma triglyceride concentrations. The c.−3G substitution forms part of the *APOA5∗2* haplotype that was previously shown to be associated with increased plasma triglyceride concentrations (25), whereas the W19 allele is part of a different haplotype, *APOA5∗3* that is independently associated with increased plasma triglyceride concentrations.

Several other polymorphisms in the apolipoprotein *A1/C3/A4/A5* gene cluster have been shown to be associated with plasma triglyceride levels in different studies, the most prominent being the *Sst*I polymorphism in the 3′ UTR of *APOC3*. In order to assess the effect of this and other polymorphisms on plasma triglyceride levels in our Caucasian cohort (Berkeley Lipid Study Population), we genotyped the *Sst*I polymorphism in *APOC3*, the −455 and −482 polymorphisms in the *APOC3* promoter and the *Xmn*I polymorphism in the 5′ flanking region of *APOA1* (for a review, see ref. 13). None of these polymorphisms showed any significant association with plasma triglyceride levels under any of the dietary conditions. This further supports the notion that the effect detected with haplotypes *APOA5∗2* and *APOA5∗3* in our study are due to sequence variants in *APOA5* and not polymorphisms in the neighboring apolipoprotein genes.

The proportion of individuals with plasma triglycerides exceeding the 90th percentile increased from 9% of S19 homozygotes to 15% of S19W heterozygotes and 37% of W19 homozygotes. Thus while each additional copy of the W19 allele significantly increased the risk of high plasma triglyceride levels, most carriers of W19 had normal plasma triglyceride levels and even homozygotes were not invariably hypertriglyceridemic. The increase in mean plasma triglyceride levels associated with the W19 allele varied among the different sexes and ethnic groups, ranging from 23 mg/dl in Caucasian women to 94 mg/dl in Caucasian men, averaging 43 mg/dl in the entire Dallas Heart Disease Prevention Project (DHDPP) sample. The effect of the *APOA5∗2* haplotype on plasma triglyceride levels also showed gender and ethnic specific effects on plasma triglyceride levels. The mechanisms underlying these gender and ethnic specific associations are not known, but the data are consistent with the notion that the two haplotypes do not produce categorical hypertriglyceridemia in most individuals, but act as modifier alleles for plasma triglyceride concentrations. Population-wide, the frequency of *APOA5∗2* and *APOA5∗3* is considerable: 35% of African-Americans, 53% of Hispanics and 24% of Caucasians carry at least one of these two minor haplotypes which have been associated with increases in plasma triglyceride levels.

The results of this study raise two important questions concerning the role of APOAV in triglyceride metabolism and atherosclerosis. First, the specific role of APOAV in plasma triglyceride transport has not been defined. Therefore the mechanisms underlying the association between the *APOA5∗2* and *APOA5∗3* haplotypes and plasma triglyceride concentrations are not known. In mice, *apoA5* deficiency is associated

**Table 4.** Ethnicity, *APOA5* genotype and plasma triglyceride concentrations in a random sample of Dallas County residents (DHDPP Population)

| | | −1131TT | −1131TC | −1131CC | *P* |
|---|---|---|---|---|---|
| African-American women | Mean ± SD | 99 ± 75 | 99 ± 76 | 128 ± 78 | 0.3 |
| | Median ± IQ range | 81 ± 73 | 85 ± 62 | 141 | |
| | *n* | (537) | (147) | (7) | |
| African-American men | Mean ± SD | 140 ± 206 | 151 ± 160 | 139 ± 125 | 0.24 |
| | Median ± IQ range | 94 ± 86 | 104 ± 95 | 85 | |
| | *n* | (376) | (106) | (5) | |
| Hispanic women | Mean ± SD | 158 ± 201 | 160 ± 91 | 270 ± 186 | 0.04 |
| | Median ± IQ range | 119 ± 78 | 140 ± 118 | 216 | |
| | *n* | (125) | (58) | (4) | |
| Hispanic men | Mean ± SD | 173 ± 150 | 250 ± 188 | 151, 430 | 0.002 |
| | Median ± IQ range | 134 ± 97 | 190 ± 233 | | |
| | *n* | (103) | (34) | (2) | |
| Caucasian women | Mean ± SD | 125 ± 98 | 135 ± 97 | | 0.39 |
| | Median ± IQ range | 103 ± 90 | 105 ± 65 | | |
| | *n* | (311) | (48) | (0) | |
| Caucasian men | Mean ± SD | 164 ± 134 | 213 ± 175 | | 0.02 |
| | Median ± IQ range | 127 ± 121 | 164 ± 183 | | |
| | *n* | (286) | (42) | (0) | |

Values are plasma triglyceride concentrations (mg/dl). For groups represented by only two individuals, the individual plasma triglyceride concentrations are given. One-sided *P*-values were calculated using Wilcoxon's test.

with increased plasma triglyceride concentrations (25), therefore we predict that *APOA5*2* and *APOA5*3* influence triglyceride levels by decreasing *APOA5* function. The *APOA5*2* haplotype does not contain any nonsynonymous substitutions and presumably acts by decreasing the amount of APOAV entering the circulation. The substitution of G for A at position −3 in a critical nucleotide of the Kozak consensus sequence may impair translation of mRNAs transcribed from this allele. In the *APOA5*3* haplotype, the substitution of tryptophan for serine at codon 19 alters the predicted signal sequence of APOAV and may result in decreased secretion of the protein, although we cannot exclude the possibility that this substitution has other effects on APOAV protein function. Alternatively, other as yet unidentified nucleotide substitutions that are in linkage disequilibrium with the W19 allele may be responsible for the association observed between the W19 allele and plasma triglyceride concentrations. Second, the role of increased plasma triglyceride concentrations in atherogenesis remains controversial. The identification of alleles that are directly associated with increased plasma triglyceride concentrations provides a means to address this question. Nordestgaard *et al.* (28) reported that a rare allele of lipoprotein lipase (G188Q) associated with increased plasma triglyceride levels was significantly more common in patients with documented coronary heart disease than in the general population. However, the frequency of the G188Q allele was less than 0.1% in the general population, therefore the effective sample size of the study was very small. The increase in plasma triglyceride levels associated with the G188Q allele (70 mg/dl) was comparable to the increase associated with the *APOA5*3* haplotype in Caucasian men in the present study. Since both the *APOA5*2* and *APOA5*3* haplotypes are relatively common, studies of the association between these haplotypes and coronary heart disease may provide insights into the role of hypertriglyceridemia in atherogenesis.

## MATERIALS AND METHODS

### Subjects

The study protocols were approved by the appropriate institutional review boards. Fasting blood samples were obtained from: i) 116 hyperlipidemic patients including 34 with Type III hyperlipidaemia, 10 with familial combined hyperlipidemia, 24 with LDL cholesterol levels exceeding the 90th percentile and 48 patients with plasma triglyceride levels exceeding 500 mg/dl; ii) 82 Caucasian men and 50 Caucasian women who were homozygous for the common allele of SNP3 (−1131T) and who had plasma triglyceride concentrations above the 90th percentile for age and sex and an equal number who were homozygous for the common allele of SNP3 (−1131T) and had plasma triglyceride concentrations below the 10th percentile for age and sex (stratified population); and iii) 2660 residents of Dallas County selected at random from census tracts who participated in the Dallas Heart Disease Prevention Project (DHDPP), a population-based study of atherosclerotic heart disease. The sample included 1392 African-Americans, 420 Hispanics and 848 Caucasians.

DNA samples were also obtained from healthy, nonsmoking, Caucasian men ($n = 354$) and women ($n = 65$) who had participated in previous dietary intervention protocols and had plasma cholesterol levels below 260 mg/dl and plasma triglyceride levels below 500 mg/dl (Berkeley Lipid Study Population) (25,29).

### DNA sequencing

The exons and flanking intron sequences of the *APOA5* gene were screened for sequence polymorphisms in 116 hyperlipidemic individuals by DNA sequencing. These individuals were only used for polymorphism discovery. DNA fragments of

**Table 5.** Oligonucleotides used in biplex Invader genotyping assays. The nucleotide highlighted in bold indicates the polymorphic base

| SNP | | Sequence |
|---|---|---|
| c.−3A>G | Probe 1 | ATG ACG TGG CAG AC**G** TAA TGG CAA GCA TGG C |
| | Probe 2 | CGC GCC GAG G**A**T AAT GGC AAG CAT GGC |
| | Invader | GCC TCC CTC CAC CTG TCT TCT CAG AGC AGT |
| c.56C>G | Probe 1 | ATG ACG TGG CAG AC**G** AAA ACG CTG TGG AGA G |
| | Probe 2 | CGC GCC GAG G**C**A AAA CGC TGT GGA GAG |
| | Invader | GCC TTT CCG TGC CTG GGT GGC CT |

~400 bp each spanning an exon and the adjacent intronic sequences or the proximal promoter sequence were PCR amplified and sequenced using BigDyeTerminator Cycle Sequencing reagents on an ABI3100 automated sequencer.

### SNP genotyping

The S19W polymorphism was assayed using PCR–RFLP and PCR Invader assays (Third Wave Technologies, Madison, WI) as described previously (Olivier *et al.*, submitted for publication). All PCR primers and probes used in biplex Invader assays for this study are listed in Table 5. As an alternative approach to assay the S19W polymorphism, oppositely-oriented oligonucleotides (AV1-F 5′ TGCTCACCTGGGCTCTGGCTCTTC and AV1-R 5′ CCAGAAGCCTTTCCGTGCCTGGG*C*GGC) were designed with a single nucleotide mismatch such that the C to G substitution that changes codon 19 from serine to tryptophan creates an *Eag*I site. PCR was performed in 20 µl volumes containing 50 mM KCl, 10 mM Tris (pH 8.3), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 1 U of Taq DNA polymerase and 200 pM of each primer. Reactions were performed in a PTC-200 Thermal cycler (MJ Research) using an initial denaturation step of 96°C for 2 min, followed with 30 cycles of 94°C for 15 sec, 70°C for 20 sec and 72°C for 30 sec. The PCR products were digested for 3 h at 37°C with 7 U of *Eag*I (New England Biolabs) in buffer provided by the manufacturer and analysed by electrophoresis in 3% agarose gels.

The −1131T>C polymorphism was analysed by mass spectrometry using the MassArray system (Sequenom Corporation) (30). The polymorphisms c.56C>G, c.−3A>G and c.457G>A are available in dbSNP under accession numbers ss4383597, ss4383596 and ss4383598, respectively.

### Statistical analysis

Statistical analyses were carried out using the SAS computer program. Plasma triglyceride concentrations were compared among different genotype groups using Wilcoxon's test. Allele frequencies were compared using Fisher's exact test. To determine pairwise linkage disequilibrium (LD) between SNPs, haplotype frequencies were estimated for 353 unrelated individuals using the Expectation-Maximization (EM) algorithm implemented in the computer program ARLEQUIN v. 2.0 (31). Predicted frequencies were compared to haplotype frequencies in grandparents of ten CEPH Utah families and an independent cohort of 171 Caucasian families. For these sample sets, haplotypes were determined using the HAPLO option of the GENEHUNTER software package (32).

## ACKNOWLEDGEMENTS

We thank H. Hobbs for thoughtful discussions, M. Basit for database support, J.-F. Cheng for technical support, R. Wilson, B. Crider, R. Cole, B. Gau and D. Savic for assistance with sequencing and genotyping, and P. Blanche, L. Holl and J. Orr for performing lipoprotein measurements. This work was supported in part by the Donald W. Reynolds Cardiovascular Clinical Research Center, the W.M. Keck Foundation, the NIH HL-53917 and HL66880 (JCC), the National Dairy Promotion and Research Board and administered in co-operation with the National Dairy Council and NIH-NHLBI Grant HL-18574 (R.M.K., E.M.R.), the NIH-NHLBI Programs for Genomic Application Grant HL66681 (E.M.R.) through the U.S. Department of Energy under contract no. DE-AC03-76SF00098 (R.M.K., E.M.R.), and an appointment to the Alexander Hollaender Distinguished Postdoctoral Fellowship Program sponsored by the U.S. Department of Energy, Office of Biological and Environmental Research, and administered by the Oak Ridge Institute for Science and Education (L.A.P.).

## REFERENCES

1. Forrester, J.S. (2001) Triglycerides: risk factor or fellow traveler? *Curr. Opin. Cardiol.*, **16**, 261–264.
2. Malloy, M.J. and Kane, J.P. (2001) A risk factor for atherosclerosis: triglyceride-rich lipoproteins. *Adv. Intern. Med.*, **47**, 111–136.
3. Connelly, P.W., Petrasovits, A., Stachenko, S., MacLean, D.R., Little, J.A. and Chockalingam, A. (1999) Prevalence of high plasma triglyceride combined with low HDL-C levels and its association with smoking, hypertension, obesity, diabetes, sedentariness and LDL-C levels in the Canadian population. Canadian Heart Health Surveys Research Group. *Can. J. Cardiol.*, **15**, 428–433.
4. Tai, E.S., Emmanuel, S.C., Chew, S.K., Tan, B.Y. and Tan, C.E. (1999) Isolated low HDL cholesterol: an insulin-resistant state only in the presence of fasting hypertriglyceridemia. *Diabetes*, **48**, 1088–1092.
5. Abney, M., McPeek, M.S. and Ober, C. (2001) Broad and narrow heritabilities of quantitative traits in a founder population. *Am. J. Hum. Genet.*, **68**, 1302–1307.
6. Perusse, L., Rice, T., Despres, J.P., Bergeron, J., Province, M.A., Gagnon, J., Leon, A.S., Rao, D.C., Skinner, J.S., Wilmore, J.H. *et al.* (1997) Familial resemblance of plasma lipids, lipoproteins and postheparin lipoprotein and hepatic lipases in the HERITAGE Family Study. *Arterioscler. Thromb. Vasc. Biol.*, **17**, 3263–3269.
7. Brenn, T. (1994) Genetic and environmental effects on coronary heart disease risk factors in northern Norway. The cardiovascular disease study in Finnmark. *Ann. Hum. Genet.*, **58**, 369–379.

8. Austin, M.A., King, M.C., Bawol, R.D., Hulley, S.B. and Friedman, G.D. (1987) Risk factors for coronary heart disease in adult female twins. Genetic heritability and shared environmental influences. *Am. J. Epidemiol.*, **125**, 308–318.
9. Heller, D.A., de Faire, U., Pedersen, N.L., Dahlen, G. and McClearn, G.E. (1993) Genetic and environmental influences on serum lipid levels in twins. *N. Engl. J. Med.*, **328**, 1150–1156.
10. Heller, D.A., Pedersen, N.L., de Faire, U. and McClearn, G.E. (1994) Genetic and environmental correlations among serum lipids and apolipoproteins in elderly twins reared together and apart. *Am. J. Hum. Genet.*, **55**, 1255–1267.
11. Talmud, P.J. and Humphries, S.E. (2001) Genetic polymorphisms, lipoproteins and coronary artery disease risk. *Curr. Opin. Lipidol.*, **12**, 405–409.
12. Busch, C.P. and Hegele, R.A. (2000) Variation of candidate genes in triglyceride metabolism. *J. Cardiovasc. Risk*, **7**, 309–315.
13. Groenendijk, M., Cantor, R.M., de Bruin, T.W. and Dallinga-Thie, G.M. (2001) The apoAI-CIII-AIV gene cluster. *Atherosclerosis*, **157**, 1–11.
14. Dammerman, M., Sandkuijl, L.A., Halaas, J.L., Chung, W. and Breslow, J.L. (1993) An apolipoprotein CIII haplotype protective against hypertriglyceridemia is specified by promoter and 3′ untranslated region polymorphisms. *Proc. Natl Acad. Sci. USA*, **90**, 4562–4566.
15. Ordovas, J.M., Civeira, F., Genest, J., Jr., Craig, S., Robbins, A.H., Meade, T., Pocovi, M., Frossard, P.M., Masharani, U., Wilson, P.W. *et al.* (1991) Restriction fragment length polymorphisms of the apolipoprotein A-I, C-III, A-IV gene locus. Relationships with lipids, apolipoproteins, and premature coronary artery disease. *Atherosclerosis*, **87**, 75–86.
16. Zeng, Q., Dammerman, M., Takada, Y., Matsunaga, A., Breslow, J.L. and Sasaki, J. (1995) An apolipoprotein CIII marker associated with hypertriglyceridemia in Caucasians also confers increased risk in a west Japanese population. *Hum. Genet.*, **95**, 371–375.
17. Shoulders, C.C., Harry, P.J., Lagrost, L., White, S.E., Shah, N.F., North, J.D., Gilligan, M., Gambert, P. and Ball, M.J. (1991) Variation at the apo AI/CIII/AIV gene complex is associated with elevated plasma levels of apo CIII. *Atherosclerosis*, **87**, 239–247.
18. Shoulders, C.C., Grantham, T.T., North, J.D., Gaspardone, A., Tomai, F., de Fazio, A., Versaci, F., Gioffre, P.A. and Cox, N.J. (1996) Hypertriglyceridemia and the apolipoprotein CIII gene locus: lack of association with the variant insulin response element in Italian school children. *Hum. Genet.*, **98**, 557–566.
19. Surguchov, A.P., Page, G.P., Smith, L., Patsch, W. and Boerwinkle, E. (1996) Polymorphic markers in apolipoprotein C-III gene flanking regions and hypertriglyceridemia. *Arterioscler. Thromb. Vasc. Biol.*, **16**, 941–947.
20. Stocks, J., Paul, H. and Galton, D. (1987) Haplotypes identified by DNA restriction-fragment-length polymorphisms in the A-1 C-III A-IV gene region and hypertriglyceridemia. *Am. J. Hum. Genet.*, **41**, 106–118.
21. Tybjaerg-Hansen, A., Nordestgaard, B.G., Gerdes, L.U., Faergeman, O. and Humphries, S.E. (1993) Genetic markers in the apo AI-CIII-AIV gene cluster for combined hyperlipidemia, hypertriglyceridemia, and predisposition to atherosclerosis. *Atherosclerosis*, **100**, 157–169.
22. Paul, H., Galton, D. and Stocks, J. (1987) DNA polymorphic patterns and haplotype arrangements of the apo A-1, apo C-III, apo A-IV gene cluster in different ethnic groups. *Hum. Genet.*, **75**, 264–268.
23. Ahn, Y.I., Valdez, R., Reddy, A.P., Cole, S.A., Weiss, K.M. and Ferrell, R.E. (1991) DNA polymorphisms of the apolipoprotein AI/CIII/AIV gene cluster influence plasma cholesterol and triglyceride levels in the Mayans of the Yucatan Peninsula, Mexico. *Hum. Hered.*, **41**, 281–289.
24. Tas, S. (1989) Strong association of a single nucleotide substitution in the 3′-untranslated region of the apolipoprotein-CIII gene with common hypertriglyceridemia in Arabs. *Clin. Chem.*, **35**, 256–259.
25. Pennacchio, L.A., Olivier, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M. and Rubin, E.M. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science*, **294**, 169–173.
26. Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283–292.
27. Kozak, M. (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell. Biol.*, **115**, 887–903.
28. Nordestgaard, B.G., Abildgaard, S., Wittrup, H.H., Steffensen, R., Jensen, G. and Tybjaerg-Hansen, A. (1997) Heterozygous lipoprotein lipase deficiency: frequency in the general population, effect on plasma lipid levels, and risk of ischemic heart disease. *Circulation*, **96**, 1737–1744.
29. Dreon, D.M., Fernstrom, H.A., Williams, P.T. and Krauss, R.M. (1997) LDL subclass patterns and lipoprotein response to a low-fat, high-carbohydrate diet in women. *Arterioscler. Thromb. Vasc. Biol.*, **17**, 707–714.
30. Buetow, K.H., Edmonson, M., MacDonald, R., Clifford, R., Yip, P., Kelley, J., Little, D.P., Strausberg, R., Koester, H., Cantor, C.R. *et al.* (2001) High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl Acad. Sci. USA*, **98**, 581–584.
31. Excoffier, L. and Slatkin, M. (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, **12**, 921–927.
32. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. and Lander, E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, **58**, 1347–1363.